

## Masked trends in field data: nematode population assessments as an example

H. Ferris <sup>a</sup> and J.-C. Prot <sup>b</sup>

<sup>a</sup> Department of Nematology, University of California, Davis, CA, USA

<sup>b</sup> Department of Plant Pathology, International Rice Research Institute, P.O. Box 933, 1099 Manila, Philippines

(Received 25 October 1991; accepted 26 October 1992)

### ABSTRACT

Ferris, H. and Prot, J.-C., 1993. Masked trends in field data: nematode population assessments as an example. *Ecol. Modelling*, 69: 243–265.

Nematode population and plant yield data from a field experiment were subjected to a range of analyses in an attempt to reveal underlying trends and relationships. Rank correlation proved to be a useful technique for preliminary indication of relationships. Several moving-average techniques were convenient, but introduced significant bias. A robust, locally-weighted regression procedure, although computationally intensive, was the most useful tool for investigating the nature of the relationships among plant-parasitic nematode populations at different densities, and between crop yield and nematode population density. In a field in which cotton and cowpea were grown in rotation, cotton yield was more strongly related to total numbers of plant-parasitic nematodes than it was to numbers of any individual species. There were indications of incompatible or antagonistic associations among the plant-parasitic nematode species, especially at higher population densities. These interactions would not have been revealed by more conventional analyses, and may provide a basis for hypothesis formulation and additional experimentation.

### INTRODUCTION

Relationships in non-time-course ecological data from field studies may be masked by variability resulting from natural variance, experimental error, and sampling error. Attempts to describe the relationships between variables by linear, quadratic, or higher-order polynomial regression have limitations:

(a) the observed trends may not be adequately described by a theoretical or statistical model;

---

Correspondence to: H. Ferris, Department of Nematology, University of California, Davis, CA 95616, USA.

- (b) the selection of the appropriate polynomial level is not immediately obvious, higher-order polynomials may reduce residuals but be of questionable biological significance;
- (c) the fit of data to the model is not flexible or local, variability measured at one extreme of the data set may strongly influence the fit in another area (Chambers et al., 1983); or,
- (d) the underlying mechanics of the relationship between two variables may change in different regions of the data set, so that a single model may be inappropriate.

Relationships between crop yield and nematode population densities are usually determined in field plots or microplots in which a range of population densities of nematodes has been established (Jones, 1956; Oostenbrink, 1966; Barker, 1985). The range of population densities is achieved by introducing known numbers of nematodes into the soil after soil fumigation, or by adjusting preexisting nematode population densities in a field site by nematicides or differential hosts. These approaches have allowed the testing of models involving single nematode species (Seinhorst, 1965) or multiple species (Duncan and Ferris, 1983). However, the environmental conditions in such studies do not fully represent whole-field situations. The approaches are designed to control sources of variation extrinsic to the model being tested, including other biotic and abiotic components. The use of nematicides to regulate nematode population densities may significantly perturb the system through effects on soil microflora and microfauna or through direct effects on crop growth (Baujard et al., 1987). Consequently, it is difficult to extrapolate observations on yield loss from field plot and microplot conditions to whole-field situations.

The synoptic or whole-field approach involves selecting a site that encompasses an acceptable range of environmental diversity, and using transects or distributing plots across the site (Ferris, 1978; Stynes et al., 1979). A larger scale alternative of this approach is the use of regional surveys (Heath et al., 1977; McKenry and Kretsch, 1987). Although controlled environment studies provide an important basis for understanding mechanisms and testing models, field studies are necessary to confirm or negate extrapolative hypotheses from microcosm and mesocosm studies. They may suggest areas for further experimentation. Conversely, without complementary reductionist studies, the presence or absence of effects in whole-field studies may be difficult to evaluate.

In both reductionist and synoptic approaches, but especially the latter, some data collapsing and smoothing is necessary to reveal underlying trends. Moving averages, although having statistical limitations, are approaches to smoothing data (Hunt, 1982). Recent applications of a smoothing procedure developed by Cleveland (1979), that does not require the

assumption of an underlying model, have considerable potential for ecological studies. The procedure was initially provided with the acronym "LOWESS" (locally-weighted scatterplot smoother), but more recently, Cleveland and Devlin (1988) have referred to the procedure as "LOESS". The technique involves running regression or "running re-fits", whereby first or second-order polynomials are fitted across restricted domains of the data set (Hunt, 1982).

The objectives of this study are (i) to evaluate some non-conventional approaches to analysis of field data, and (ii) to use those approaches to investigate relationships between yield and nematode populations, and among co-inhabiting plant-parasitic nematode species, using sample data from cultivated fields.

## MATERIALS AND METHODS

### Data

Soil samples were collected from a field located near Pixley in Tulare County, California, in which cowpea (*Vigna unguiculata* cv. California Blackeye 5) and cotton (*Gossypium hirsutum* cv. Acala SJ2) were grown in rotation. The field was planted to cowpea in 1982, held fallow during the winter of 1982-83, and planted to cotton in 1983. A more detailed description of the field site is provided in Ferris et al. (1990).

A 25 × 25 grid pattern was superimposed on the field, with grid intersection points 6.5 m apart. The field was sampled 11 times between September 1982 and October 1983; data from four of those sampling dates are used in the present study: September 1982 at harvest of the cowpea crop, March 1983 at planting of the cotton crop, June 1983 at midseason of the cotton crop, and October 1983 at cotton harvest. At each sampling date, an individual core of soil (2.5-cm diam, 30-cm depth) was removed from within a circle of 30-cm radius at each grid intersection. In October 1983, the weight of cotton seed and fiber, expressed in g/plant, was determined for plants within 30 cm of each grid intersection.

Each soil core was weighed. Nematodes were extracted from the individual soil cores using a semi-automatic elutriator and sugar flotation (Byrd et al., 1976). Numbers of individuals of each species of plant-parasitic nematodes were recorded per 100 g soil for each core. Plant-parasitic nematode species present in the cores included *Criconemella xenoplax* (Luc and Raski), *Helicotylenchus dihystrera* (Sher), *Paratrichodorus minor* (Siddiqi), *Meloidogyne incognita* (Chitwood), *M. javanica* (Chitwood), *Pratylenchus* spp. (Filipjev), and *Xiphinema americanum* (Cobb). Because *C. xenoplax*

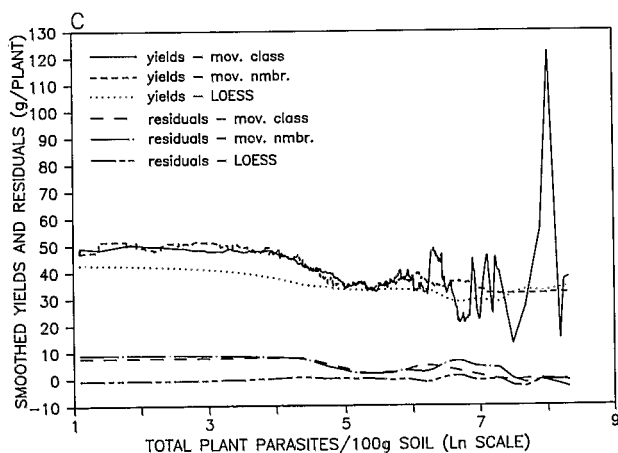
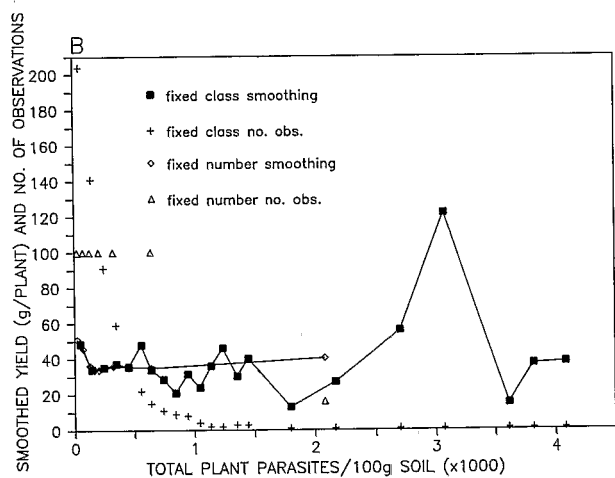
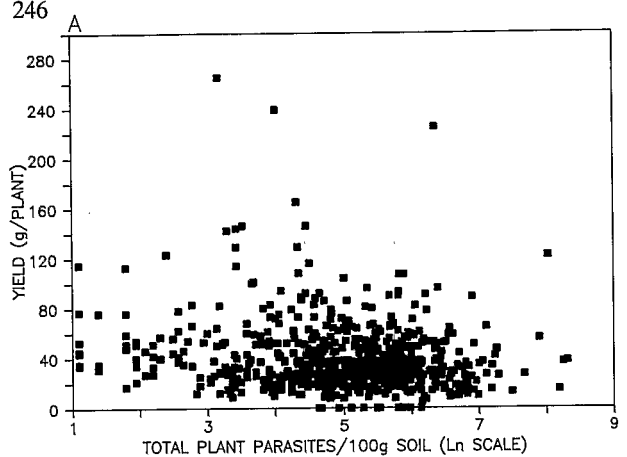


Fig. 1. Nature of the data set of cotton yield in October 1983 and total numbers of plant-parasitic nematodes in September 1982, and the result of various smoothing techniques. (A). Unsmoothed data. (B). Fixed-class and fixed-number smoothing techniques. (C). Moving-class, moving-number, and LOESS techniques.

and *Pratylenchus* spp. were present in low numbers in the samples, they were not considered for this study.

The data set encompasses great variability because the nematode count data are based on single cores of soil, and yield data are based on single plants (Fig. 1A). The data set lacks any smoothing effect that would be produced if the soil cores were composited. We considered the merits and constraints of a series of approaches designed to reveal the underlying trends of the data set.

## Analyses

### *Spearman's rank correlation*

Ranked correlation procedures are appropriately used when standard correlation procedures are not valid because neither variable is normally distributed (Snedecor and Cochran, 1968). Spearman's method (Spearman, 1904) involves ranking both variables and calculating the ordinary correlation coefficient of the rankings. The rank correlation coefficient,  $r_s$ , ranges from  $-1$  (indicating complete discordance in the data) through  $0$  (indicating that the variables vary independently, i.e., the null hypothesis) to  $1$  (indicating complete concordance).

On the four sampling dates considered in this study, the relationship between cotton yield and total numbers of plant-parasitic nematodes in the samples was analyzed by Spearman's rank correlation. These analyses were also conducted for the relationships between cotton yield and population levels of *H. dihystra*, *Meloidogyne* spp., *P. minor*, and *X. americanum* considered separately. Numbers of individuals of *H. dihystra*, *P. minor*, and *Meloidogyne* spp. were compared in pairs. Coefficients were determined using all 625 cores from the field, or only those cores in which one of the species considered was represented by at least 10, 25, 50, 75, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 individuals. Numbers of plant-parasitic nematodes recovered during the winter months, and during the cotton season, were generally low. Consequently, the different species were compared only for September 1982 (just after cowpea harvest) and October 1983 (just after cotton harvest).

### *Ranking and fixed-class averages*

This approach was used by Ferris (1984) to determine appropriate model structure and parameter values for relating crop yields to preplant, single-species nematode population densities. A series of  $j \times y$  observations, ranked on the magnitude of the  $x$  values, is divided into classes of equal size,  $d$ ; the  $\mu_x$  and  $\mu_y$  values are computed, such that  $\mu_x = \sum x_i / n_k$  and  $\mu_y = \sum y_i / n_k$ , where  $n$  is the number of observations in the  $k$ th class. The

number of reduced points in this approach is equal to the number of classes selected, and may be much smaller than the original scope of the data set. The classes are equal in size, but each class contains different numbers of observations, so the weight ascribed to each class differs. The range of the smoothed data set is less than that of the original, as it extends from average point in the first class to the average point in the last class, rather than from the first ranked observation to the last ranked observation.

#### *Ranking and moving-class averages*

For a series of  $j$   $x, y$  observations, ranked on the magnitude of the  $x$  values, the  $\mu_y$  values are computed such that  $\mu_y = \Sigma y_i / n_k$ , where  $n$  is the number of observations in the  $k$ th class. Each class ranges from  $(x_i - d/2)$  to  $(x_i + d/2)$  where  $d$  is the selected size of classes. If  $(x_i - x_1) < d/2$ , the class ranges from  $x_1$  to  $(x_i + d/2)$ ; if  $(x_j - x_i) < d/2$ , the class ranges from  $(x_i - d/2)$  to  $x_j$ . The number of smoothed points is equal to the number of observations. Classes are equal in size, but differ in the number of observations. Thus, each smoothed observation has different weight in any analysis. This problem can be overcome by using the smoothed data set to suggest the appropriate model to describe the data, and obtaining parameter values for that model from the original data set. The smoothed points at the extremes of the ranked data set may be less robust, as they are based on skewed classes. For example, the first smoothed mean is based only on observations within  $d/2$  greater than  $x_i$ , and the last is based only on observations  $d/2$  less than  $x_j$ . The process is computationally complex, especially when spreadsheet procedures are used, because calculations require repeated looping and conditional checks to determine whether each data point belongs in the current class.

#### *Ranking and fixed-number averages*

A series of  $j$   $x, y$  observations, ranked on the magnitude of the  $x$  values, is divided into classes of equal numbers of observations,  $n$ . For each class, the  $\mu_x$  and  $\mu_y$  values are computed, such that  $\mu_x = \Sigma x_i / n$  and  $\mu_y = \Sigma y_i / n$ . The number of reduced points is a function of the number of times the total observations can be subdivided into the specified observation groupings. There are equal numbers of observations in each interval, but the class size differs. Unlike ranking and moving-class averages, this approach is computationally easy using spreadsheet procedures. As with fixed-class averages, the range of the smoothed data set is less than that of the original, as it extends from the average point of the first  $k$  set of observations to the average point of the last  $k$  set of observations, rather than from the first ranked observation to the last ranked observation.

### *Ranking and moving-number averages*

For a series of  $j$   $x, y$  observations, ranked on the magnitude of the  $x$  values, the  $\mu_y$  values are computed, such that  $\mu_y = \sum y_i / n$ , where  $n$  is the number of observations selected as an interval grouping. Each interval ranges from  $x_{(i-n/2)}$  to  $x_{(i+n/2)}$ . If  $(i - n/2) < 1$ , the interval ranges from  $x_1$  to  $x_{(i+n/2)}$ ; if  $(i + n/2) > j$ , the interval ranges from  $x_{(i-n/2)}$  to  $x_j$ . The number of smoothed points is equal to the number of observations in the data set. The points used for each smoothed mean value are equally distributed on either side of the mean. At each extreme of the data set, only half the number of points are considered, that is, the first  $n/2$  points or the last  $n/2$  points. Within the bounds of the first and last  $n/2$  points, there are equal numbers of observations per interval, but class sizes differ. Like ranking and fixed-number averages, this approach is computationally easy using spreadsheet procedures; however, standard errors for the smoothed values are difficult to compute and interpret. The appropriate procedure is, again, to select the model from the smoothed points, and to use the original data set for regression parameters and standard error calculations.

### *Application of averaging techniques*

The relationship between cotton yield in October 1983 and total nematode population levels in September 1982, March 1983, June 1983, and October 1983 was examined by fixed- and moving-class averages (class sizes of 100 as an example of those tested), and by fixed- and moving-number averages (100 observations in each class as an example of those tested). Residuals (smoothed minus actual yield levels) were calculated for each of the population data points. The residuals were subjected to the smoothing effect of the LOESS technique to reveal any localized bias attributable to either moving-class or moving-number approaches, as suggested by Cleveland (1979). The sums of squares of residuals for the whole data set were calculated for each of the moving-average techniques as a measure of overall goodness-of-fit of the smoothed values.

### *LOESS*

This procedure combines the strengths of both the moving-class and moving-number procedures, and elegantly compensates for their limitations. The  $x, y$  data set is ranked according to the magnitude of the  $x$  values, and a proportion,  $d$ , of the data points to be included in each average is selected (usually between 0.5 and 0.8). In this regard, LOESS is similar to the moving-number approach. An iterative smoothing procedure is invoked. In the first iteration, a weight function,  $W_i$ , is selected such that at the  $x_i$  for which the average is being calculated,  $W_i = 1.0$ , and at the  $d$ th

nearest data point to  $x_i$ ,  $W_i = 0.0$ . The  $W_i$  function is symmetrical on each side of  $x_i$ . Weighted, least squares procedures are used to fit a  $n$ th order polynomial (we used  $n = 1$ ) to the weighted data points within the selected range. Each smoothed data point is calculated from the regression equation at  $x_i$ . This procedure is repeated for  $i = 1$  to  $m$ , where  $m$  is the total number of observations. At the extremes of the data set, the same proportion of observations is still used, its boundaries constrained by the first and last data set value. In a second iteration, the process is repeated but, in this case, the effect of the data point in the regression is weighted not only by the  $W_i$  function, but also by a weight that reflects the size of the residual of the data point from the previous fitted line. Here, a tricubed function is used to calculate a weight,  $B_i$ , such that  $0 \leq B_i \leq 1.0$ . Each of the data points used to determine the smoothed line at  $x_i$  is subjected to weighted, least-squares regression, using the weight  $W_i B_i$ . This final, dual-weighted iteration is repeated at least twice. The procedure is detailed in Cleveland (1979, 1985), Chambers et al. (1983), and Cleveland and Devlin (1988).

In the LOESS procedure, each smoothed point is based upon the same number of observations, providing the benefit of the moving-number approach, while also minimizing biases associated with different class sizes by weighting the contributions of the points to the least squares regression fit according to their distance from the pivotal point of the class. In addition, LOESS introduces a robustness feature by considering the distance of each point in the data set from the central trend in that region by weighting the contribution of the points relative to their distribution about the central trend. Thus, the procedure results in a robust fit of a line that is locally sensitive. The tricubed weighting feature, that weights each point according to the magnitude of its residual from the central trend, assumes that the residuals are symmetrically distributed about the trend. This assumption may be invalid when data points are bounded by zero at the lower extreme and unbounded at the upper extreme. In our analyses, asymmetries were reduced by our elimination of zero values from the data set. We feel that violation of the symmetry assumption is offset by the advantages provided by the robustness procedure in revealing trends and associations in the data set. One disadvantage of the LOESS procedure is that it is computationally intensive (Cleveland, 1979, 1985; Chambers et al., 1983). In our studies, a data set of 625 pairs of observations required approximately 5 h of computation on an IBM PC/AT-compatible desktop computer with an 80286 processor and a math co-processor.

The relationship between yield of cotton in October 1983 and total nematode population levels in September 1982, March 1983, June 1983, and October 1983 was examined by LOESS for comparison with similar analyses using averaging techniques. Since we were interested in the impact



of plant-parasitic nematodes on cotton yield, we did not include coordinates for which zero yield was recorded. On some dates, nematodes of an individual species were not detected in a substantial number of cores. Consequently, we used only those cores in which the species of interest was detected, since yield may be influenced by many other factors at the other sites. We also used LOESS to examine relationships between total nematode population data sets, and between population levels of key plant-parasitic nematodes on each of the selected sampling dates. In examining relationships between individual species, we omitted all samples in which either species was absent. We assumed that the species could interact only if both were present, and that the absence of one was not necessarily an indication of a direct effect of the other.

In the LOESS analyses, we used half of the points to calculate each smoothed value, and two smoothing iterations were conducted after calculation of the initial set of residuals. Residuals (smoothed minus actual yield levels) were calculated for each of the population data points. The residuals were subjected to the smoothing effect of the LOESS technique to reveal any localized bias attributable to the LOESS procedure, and also for comparison with similar analyses for either moving-class or moving-number approaches (Cleveland, 1979). The sums of squares of residuals for the whole data set were also calculated for the LOESS technique.

## RESULTS

### *Comparison of smoothing techniques*

The sums of squares of residuals (SSRs) were lowest for the moving-class technique ( $464.2 \times 10^3$ ), followed by the moving-number technique ( $476.9 \times 10^3$ ) and, finally, LOESS ( $495.4 \times 10^3$ ), indicating overall goodness-of-fit of the smoothed points to the original data achieved by each technique. By contrast, when a single (linear) regression model was used for the whole data set, the SSR was  $495.2 \times 10^3$ . The magnitude of the SSRs does not address the issue of local bias, nor of differences in the underlying relationship in different regions of the data set, associated with the use of each technique. The sum of the smoothed residual values for each smoothing method provides a measure of the cumulative bias of the smoothing techniques across the data set. The sum of smoothed residuals was 3057.4 for the moving-class technique, 3089.1 for the moving-number technique, 76.9 for LOESS, and 2999.3 for the linear regression, fixed model.

### *Ranking and fixed-class averages*

The distribution of observations across the range of fixed classes was highly skewed. The lowest class contained 209 observations, while the

classes at higher population densities contained only one observation each (Fig. 1B).

#### *Ranking and moving-class averages*

Although there were many observations available for each class average at low population densities (generally more than 100 observations for each class increment), there were fewer observations at high population densities. The last seven classes contained only one observation each, resulting in no smoothing effect for those classes (Fig. 1C). There was a tendency for the smoothed values to overestimate the yield trend at low population densities, as evidenced by the smoothed residuals. Because of the low number of observations available for each class at higher population densities, smoothed residuals were low since there was little or no deviation between actual and smoothed yield values in that region of the data set (Fig. 1C).

#### *Ranking and fixed-number averages*

There were only six classes across the data set when class-sizes of 100 observations were used. The classes were aggregated in the lower population density range of the data set and were less frequent at higher population densities (Fig. 1B).

#### *Ranking and moving-number averages*

A smooth line developed, but classes were of different sizes, and each observation had the same weight in its contribution to the smoothed mean (Fig. 1C). The smoothed values tended to overestimate the yield trend at low population densities, as evidenced by the smoothed residuals. Since there were equal numbers of observations in most classes, decreasing to no less than half of the moving-number class size at extremes of the ranked data set, smoothed values were less erratic at high population levels than for the moving-class approach; class sizes at the upper extreme of the data set involved very few observations. Again, there was less evidence of localized bias in the upper region of the data set (Fig. 1C).

#### *LOESS*

Down-weighting observations at greater distances from the points in question, and down-weighting "outliers" about the smoothed line, resulted in damage estimates considerably lower than those estimated by moving average techniques (Fig. 1C). There was also an indication that the nature of the relationship changed at higher population densities, but this observation is less reliable as the spread of the data was greater in the upper regions of the data set. In comparison to the moving-class and moving-

TABLE 1

Spearman's rank correlation coefficients between yield and total nematode population, and between yield and each nematode species, at each sampling date

	September 1982	March 1983	June 1983	October 1983
Total population	-0.2314 **	-0.2146 **	-0.2508 **	-0.3255 **
<i>Meloidogyne</i> spp.	-0.2419 **	-0.1542 **	-0.0859 *	-0.2276 **
<i>Helicotylenchus dihystera</i>	-0.0961 *	-0.1100 **	-0.1585 **	-0.2396 **
<i>Paratrichodorus minor</i>	-0.0525	-0.2026 **	-0.2452 **	-0.1594 **
<i>Xiphinema americanum</i>	-0.1555 **	-0.0967 *	-0.1458 **	-0.1754 **

Levels of significance: \* = 5%, \*\* = 1%.

number averaging techniques, the smoothed yield values were not locally biased at low population densities, as evidenced by the smoothed residuals. The smoothed residuals also showed less bias at higher population densities than did those of the other techniques (Fig. 1C).

Based upon our evaluation of smoothing techniques, we eliminated fixed-class and fixed-number averaging from further consideration. We concentrated on using the LOESS approach for data set smoothing, with consideration of moving-class and moving-number techniques where appropriate.

### *Nematode-yield and nematode-nematode relationships*

#### *Spearman's rank correlation*

There was a significant negative Spearman's rank correlation coefficient between yield and numbers of total plant-parasitic nematodes, and between yields and numbers of individual nematode species for each sampling date (Table 1). On each date considered, the strongest correlation was invariably between yield and the total plant-parasitic nematode population. Of the four plant-parasitic nematode species considered separately, *Meloidogyne* spp. was best correlated with plant yield in September 1982, and *P. minor* in March, June, and October 1983 (Table 1).

*Paratrichodorus minor* was omnipresent in the field; in the soil cores collected from the 625 sampling sites, the nematode was detected in 532 soil cores in September 1982, and in 624 cores in October 1983 (Table 2). *Meloidogyne* spp. and *H. dihystera* displayed similar dispersion patterns. In September 1982, juveniles of *Meloidogyne* spp. were not detected at 82

TABLE 2

Numbers of detections and average numbers (per 100 g soil) of *Helicotylenchus dihystera*, *Paratrichodorus minor*, and *Meloidogyne* spp. juveniles per sample site in September 1982, after cowpea harvest, and in October 1983, after cotton harvest

	<i>Meloidogyne</i> spp.	<i>H. dihystera</i>	<i>P. minor</i>
Numbers of detections			
September 1982	543	436	532
October 1983	397	514	624
Average numbers of nematodes/site			
September 1982	183	120	28
October 1983	166	355	193

sites, and *H. dihystera* was not detected at 189 sites; 48 of the 189 sites coincided with areas where *Meloidogyne* spp. were not detected. Spearman's rank correlation was applied to all cores in which at least a threshold number of individuals of either species was present; those cores not meeting this constraint were eliminated from consideration. The threshold number of individuals was increased progressively in the series 10, 25, 50, 75, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, and 1000, in sequential analyses. In all comparisons, the correlation coefficient between the species was positive at population values of zero individuals, and became negative when comparisons were restricted to samples with at least low populations of one species present (Fig. 2). At higher population densities of either of the species, the relationship between *Meloidogyne* spp. and *H. dihystera*, in both September 1982 and October 1983, was consistently negative. In October 1983, *Meloidogyne* spp. and *P. minor* were positively correlated at high population densities of either species. The relationship between *H. dihystera* and *P. minor*, in October 1983, appeared neutral when the population densities of either species were high.

#### *Averaging techniques*

Moving-class and moving-number averages, when plotted on a natural log scale, revealed a general relationship between cotton yield, in October 1983, and total plant-parasitic nematode population, in September 1982, in the form of a "Seinhorst equation" (Seinhorst, 1965):

$$y = m + (1 - m)z^{(p-t)},$$

where  $y$  is the relative yield,  $m$  the minimum yield at highest nematode population levels,  $z$  a damage rate parameter,  $p$  the population level, and  $t$  the tolerance level below which no yield loss is observable. For both smoothing techniques, the tolerance level was around 55 total nematodes/

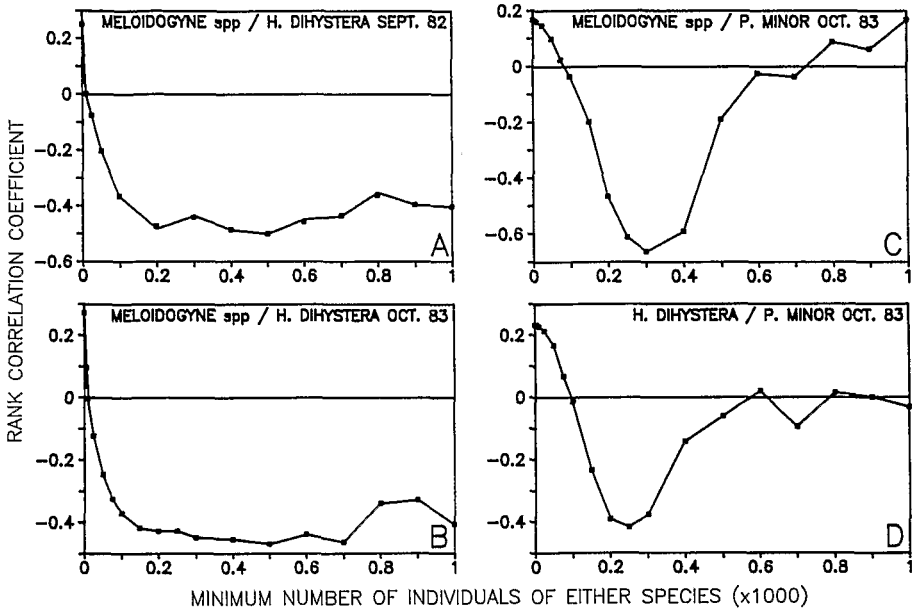


Fig. 2. Rank correlation coefficients between pairs of plant-parasitic nematode species where numbers of individuals of either of the species is above a prescribed level.

100 g soil, and the apparent minimum yield was about 0.65, most reliably determined by the moving-number approach (Fig. 1C). The maximum yield, at lowest nematode population levels, was approximately 50 g/plant.

### LOESS

Although LOESS procedures were performed on non-transformed yield and population data, the relationships are most clearly revealed by plotting the smoothed lines on a log-transformed scale of the population values (Fig. 3). For the relationship between crop yield in October 1983 and total nematode population in September 1982, with recognition and down-weighting of outliers, the LOESS-smoothed line falls below those of the moving-class average methods as nematode population densities increase (Fig. 1C). With zeros omitted, there were 604 observations remaining in the data set. As with the moving-average methods, a Seinhorst equation was revealed, with a tolerance level of approximately 20, or  $\ln 3$ , nematodes/100 g soil (Fig. 3), generally shifted left from the moving-average curves (Fig. 1C), and a minimum yield,  $m$ , of 0.7 reached at about 800, or  $\ln 6.7$ , nematodes/100 g soil. Yield values were slightly higher associated with the highest population levels, possibly suggesting reversal of causality in those plants. However, at the highest population levels, yields were still 20%

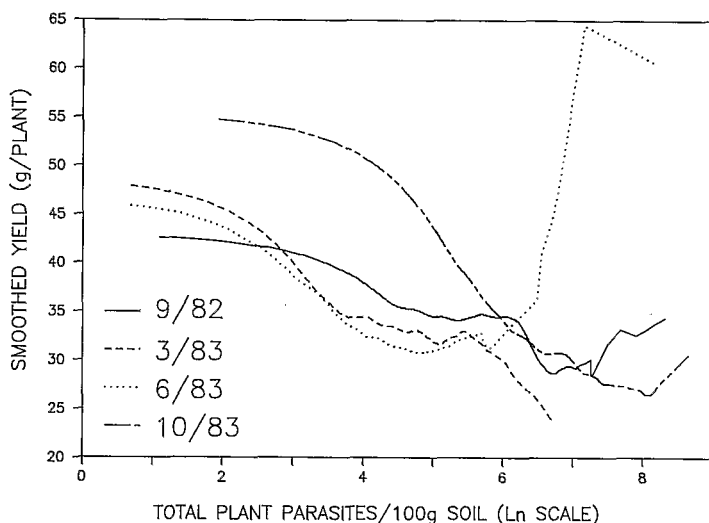


Fig. 3. LOESS plots of the relationship between cotton yield and total numbers of plant-parasitic nematodes at four sampling dates (September 1982, March 1983, June 1983, and October 1983).

lower than the maximum yield levels of about 43 g/plant associated with lowest population levels (Fig. 3).

There were 599 observations remaining in the data set of cotton yields for October 1983 and total plant-parasitic nematode populations for March 1983, with zeros omitted. Yield decreased at higher total plant-parasitic nematode population levels. Yields associated with highest population levels were 50% ( $m = 0.5$ ) lower than those associated with lowest levels, and a tolerance level of approximately 7.4, or  $\ln 2$ , was suggested (Fig. 3). There were 579 non-zero observations in the data set of cotton yield values for October 1983 and total plant-parasitic nematode numbers for June 1983. Yields were lower as population levels increased initially, but high yields were associated with high nematode population levels (Fig. 3). Application of the Seinhorst model to yields associated with the smoothed June 1983 total nematode population data seems inappropriate. There were 612 non-zero observations in the data set representing yield and total plant-parasitic nematode numbers at cotton harvest in October 1983. Yields decreased as nematode population levels increased; for example, they were 50% lower associated with population levels of about 2500 nematodes/100 g soil as compared with yields associated with low population levels. There was an indication of slightly higher yields associated with the highest population levels (Fig. 3). Because the Seinhorst model implies a cause and effect relationship, and is usually applied to the relationship

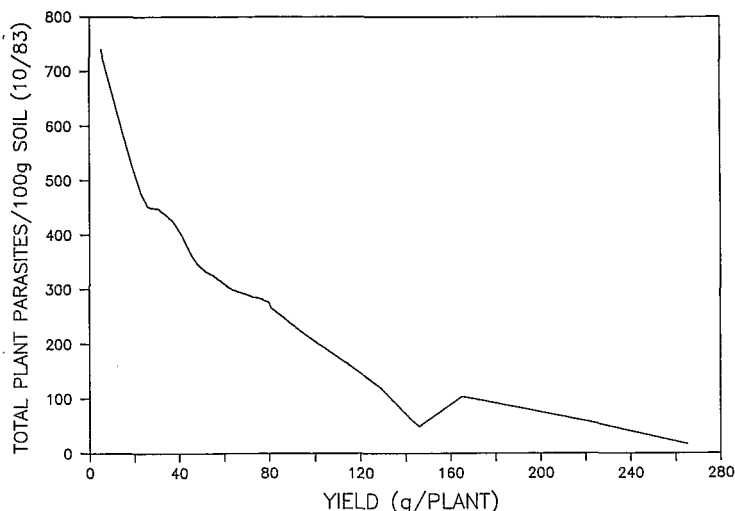


Fig. 4. LOESS plot of the relationship between total numbers of plant-parasitic nematodes and cotton yield in October 1983.

between yield and nematode population levels prior to planting, or early in the growing season, its application to these data is inappropriate.

Since the population levels in October 1983 might be a reflection of the growth and vigor of the cotton plants constituting the food source of the nematodes, it seemed reasonable to conduct a reciprocal analysis of the relationship between total population density and plant yield. Population levels were highest associated with low yields per plant, and lowest associated with high yields per plant, indicating that, for this field, the population levels of plant-parasitic nematodes were not positively influenced by plant growth (Fig. 4). However, plant growth and vigor may not be directly related to yield of seed and lint. In many crops, fruit yields may be lower on vegetatively vigorous plants.

In the 532 non-zero samples of *Meloidogyne* spp. measured in September 1982 and cotton yield in October 1983, yield values associated with highest nematode levels were only about 40% of those at lowest nematode levels (Fig. 5A). We continue to assume that the population measured in September 1982 is predominantly *M. javanica*, although limited reproduction of *M. incognita* occurs on cowpeas (Duncan and Ferris, 1983). An obvious question is whether or not the *Meloidogyne* spp. population measured in October 1983 is related to that measured in September 1982, in terms of numbers and spatial pattern. There was no definite trend in the 324 non-zero observations of *Meloidogyne* spp. measured in March 1983 on yield measured in October 1983 until the population level reached approxi-

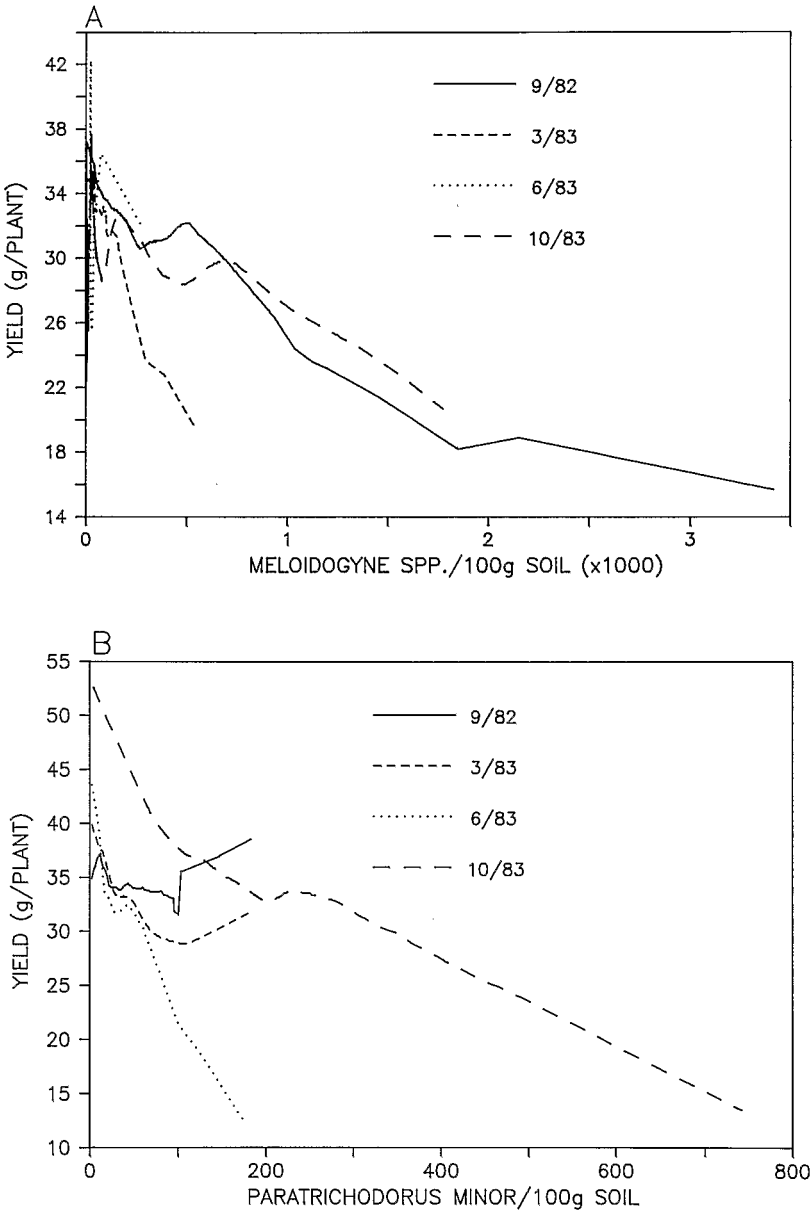


Fig. 5. LOESS plots of the relationships between cotton yield in October 1983 and numbers of individual plant-parasitic nematode species in September 1982, March 1983, June 1983, and October 1983. (A). *Meloidogyne* spp. (B). *Paratrichodorus minor*.

mately 60 juveniles/100 g soil. Above this level, yields were about 40% lower when the highest population levels were reached (Fig. 5A). It is likely that many of the juveniles in the samples were *M. javanica* from the



preceding cowpea crop. Population levels were low or non-detectable at many sample locations. In June 1983, population levels of *Meloidogyne* spp. were very low and infrequent, with non-zero observations detected in only 140 samples. There was no apparent relationship between yield and population levels; the data were therefore excluded from Fig. 5A. In the 393 non-zero observations of *Meloidogyne* spp. measured in October 1983 and yield, measured in the same month, cotton yields were 42% lower associated with highest population levels than with lowest levels (Fig. 5A).

There was no apparent relationship between cotton yield measured in October 1983 and population densities of *P. minor* measured in September 1982 (522 non-zero observations) (Fig. 5B), confirming the indications provided by Spearman's rank correlation. Smoothed yield values corresponding to the highest population levels were the same, or slightly higher, than those observed for the lowest population levels. The relationship between yield measured in October 1983 and *P. minor* population densities measured in March 1983 (565 non-zero observations) was negative and almost linear, except at the highest population levels (Fig. 5B). Smoothed cotton yield values associated with highest levels of *P. minor* were about 30% lower than those associated with the lowest population levels. In the 550 non-zero observations of *P. minor* measured in June 1983 and cotton yield measured in October 1983, the relationship was linear and strongly negative, with yields associated with highest population levels approximately 70% lower than those associated with the lowest levels (Fig. 5B). Cotton yields in October 1983 tended to be lower on those plants with high densities of *P. minor* measured in October 1983 (611 non-zero observations). The relationship was almost linear, with yields associated with high population densities less than 25% of those observed at the lowest population densities (Fig. 5B).

Due to the apparent association of yield with the total nematode population measured in September 1982, and the association of yield and numbers of *Meloidogyne* spp. and *P. minor* measured in October 1983, we examined the relationship between numbers of each of these species measured in October 1983 and total population measured in September 1982. Numbers of *P. minor* measured in October 1983 showed little or no association with total plant-parasitic nematode population levels measured in September 1982 (Fig. 6A). Numbers of *M. incognita* measured in October 1983 increased in association with total nematode population levels recorded in September 1982 (Fig. 6A). The October 1983 levels of *M. incognita*, associated with total population levels of plant parasites in September 1982 of 2000 nematodes/100 g soil, were approximately double those associated with lowest total plant-parasite levels. The *M. incognita* population levels in October 1983 were considerably higher associated with

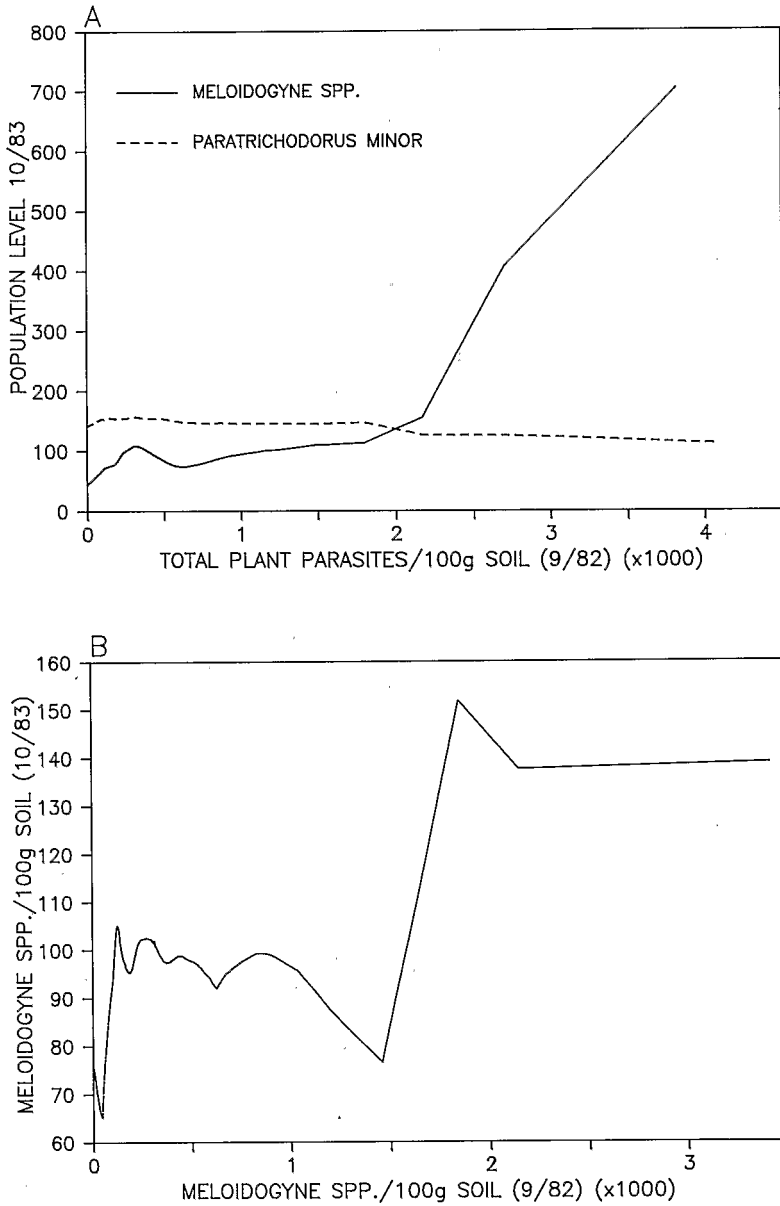


Fig. 6. LOESS plots of the relationship between numbers of plant-parasitic nematodes in October 1983 and September 1982. (A). *Meloidogyne* spp. and *Paratrichodorus minor* in relation to total numbers of plant-parasitic nematodes. (B). Relationship between numbers of individuals of *Meloidogyne* spp. on the two dates.

the few samples with more than 2000 total plant parasites/100 g soil, measured in September 1982 (Fig. 6A).

Due to the apparent relationship between yield decline observed in October 1983 and high population levels of *Meloidogyne* spp. in September 1982, it seemed instructive to examine the relationship between populations of *Meloidogyne* spp. measured in October 1983 and in September 1982 (Fig. 6B). Since we assume that the October 1983 population was predominantly *M. incognita*, due to the host status of cotton for this species, we would expect that, if there were significant numbers of *M. incognita* present in September 1982, these sampling sites would also contain high populations in October 1983. Such a relationship would explain the apparent decline in cotton yield at high population levels observed in September 1982. There were 363 sampling sites in which *Meloidogyne* spp. was detected in both September 1982 and October 1983. It appears that very low populations of presumed *M. incognita* in October 1983 were associated with very low populations of *Meloidogyne* spp. measured in September 1982. There was no discernible relationship between October 1983 population levels and September 1982 levels in the range of approximately 75 to 750 juveniles/100 g soil. Above that level, the highest population levels of *M. incognita* were associated with the highest population levels of *Meloidogyne* spp. measured in September 1982.

## DISCUSSION

In ecological data derived from field observations or experiments, there may be masked information that provides insights and bases for hypotheses that are testable in controlled experiments in microcosm. There are myriad analyses that can be conducted by applying techniques of the type used in this study to field data sets. The challenge, then, is to develop hypotheses and explanations of the relationships that are suggested. We have elected not to be exhaustive in the analyses but, rather, to introduce and explore the utility of these interpretive and investigative approaches. Certainly there is a danger that application of smoothing techniques may lead to spurious conclusions, but we concur with Hunt (1982) that these powerful techniques have been underutilized for examination of large data sets. We emphasize the importance of verification of the indications from such preliminary analyses.

Spearman's rank correlation allows a simple approach to search for underlying associations, relationships, and interactions in a data set. The creation of data classes by various techniques provides some insights, but may be of limited value due to inherent biases. The LOESS technique (Cleveland, 1979) elegantly addresses and compensates for those biases,

although it is computationally intensive. More recent formulations and uses of LOESS (Cleveland and Devlin, 1988) involve less computation, while retaining the fundamental power of the approach.

Because of the skewed nature of nematode population data sets, with variance greater than mean, and disproportionately higher numbers of samples with low population counts, fixed- and moving-classes representing high population levels contained few observations. In trial analyses with fixed- and moving-class averages, we transformed the data set to a natural log scale. This provided greater uniformity in numbers of observations in the classes representing higher population densities, and is appealing in consideration of the decreased damage per plant per individual as nematode population densities increased. (Seinhorst, 1965; Ferris, 1984). However, we noted that log transformation skewed the data sets in the other direction and minimized robustness at low nematode population densities. Square-root transformation was a little, but not much, better. It seems unwise to introduce a transformation and to assume that it is improving characteristics of the data set because of a seductively appealing result. In this case, the transformation introduced a set of biases that were a mirror-image to those that we were attempting to eliminate. We also conducted trial analyses with the LOESS procedure on log-transformed data. We hoped to avoid the situation, at higher population densities, in which smoothed values are substantially influenced by observed values and residuals far removed from the  $x$ -value for which they are being calculated. We finally rejected this approach for the same reasons encountered in the fixed- and moving-class methods. Rather, we selected  $d$  so that the residuals showed no bias, as advocated by Cleveland (1979). We conclude that the most appropriate approach in these analyses is to accept the strengths and weaknesses of the data set. If analytical biases are introduced by distribution of the data, they should be recognized and explored.

Both Spearman's rank correlation and the various smoothing techniques revealed a negative correlation between yield of individual cotton plants and the total number of plant-parasitic nematodes in soil samples associated with those plants. Of course, conclusions regarding causality cannot be made or implied from these analyses. Rather, associations are revealed or suggested that require further experimentation to allow interpretation. In some cases, the association may be fortuitous and, in others, it may be mediated by a third and unmeasured factor such as soil texture, previous host growth patterns, or the distribution of another pest species. However, the plant-parasitic nematodes may be directly responsible for the observed depression of yield; controlled experiments in microplots will be necessary to investigate various hypotheses of causality.

If plant-parasitic nematodes are the direct cause of suppressed cotton

yield, the stronger correlation between yield and the total numbers of plant-parasites, rather than between yield and the individual species, suggests a synergism of effects. It may also suggest that all four species are affecting yield, but that their effects differ spatially across the field. The potential contribution of all the species to yield loss underscores the value of a synoptic approach (Stynes et al., 1979) to unravelling the causality of yield loss. *Meloidogyne incognita* is a well-documented pathogen of cotton (Sasser, 1972); *H. dihystra* has been associated with minor yield reduction (Bernard and Hussey, 1979); and no information is available for the affect of *X. americanum*. The presence of *P. minor* has little effect on cotton growth in greenhouse and microplot experiments, but growth reduction has been observed in association with high population levels (Sasser, 1972).

If *P. minor* has little influence on cotton yield, why does it have the strongest negative correlation with yield on three of the sampling dates? Perhaps it constitutes an indicator of the structure and pathogenicity of the total nematode community. At a particular sampling date, *P. minor* may be the species that best represents the total numbers of plant-parasitic nematodes present. For example, the extraction techniques used in these studies detect only active, vermiform stages of nematodes in the soil. The number of *Meloidogyne* spp. juveniles recovered in March 1983 does not adequately represent the pathogenic potential of these nematodes. In March, at the end of the winter season, the surviving *Meloidogyne* spp. may be predominantly in the egg stage while, in June, the population is predominantly in parasitic stages within roots. The poor relationship between yield in October 1983 and *Meloidogyne* spp. in March 1983 and June 1983 is also revealed by the LOESS smoothing procedure.

We conclude that the observed relationships between the various nematode species is unlikely to be coincidental. If the species have different environmental requirements, one species might be found in areas where the other is absent. This was not the case, as *P. minor* was omnipresent in the field and, in most cases, *H. dihystra* was absent when *Meloidogyne* spp. was also absent. The numbers of *Meloidogyne* spp. were higher when the numbers of *H. dihystra* were higher. When the three nematode species were compared in pairs, there was a strong positive correlation between numbers of individuals at low population densities. These nematodes appear to share environments of similar conditions; the more favorable the environment is to one species, the more favorable it is to the remaining two species.

Obviously, hypotheses on interactions of coinhabitant nematode species and the effects of those species on plant yield, suggested by our analyses, must be evaluated in controlled, well-designed experiments. Better understanding of nematode-nematode relationships under field conditions may

be important from the perspective of management tactics such as biological control, cropping sequences, or resistant cultivars. Any of these species-specific tactics will disrupt the equilibrium of the nematode community; the control of one predominant species may remove limitations on the multiplication of another species.

#### ACKNOWLEDGEMENTS

The authors acknowledge the suggestions and advice of Dr. Donald R. Strong, Bodega Marine Laboratory, University of California, regarding locally-weighted regression techniques.

#### REFERENCES

- Barker, K.R., 1985. The application of microplot techniques in nematological research. In: K.R. Barker, C.C. Carter and J.N. Sasser (Editors), *An Advanced Treatise on Meloidogyne*, Vol. 2, Methodology. North Carolina State University Graphics, Raleigh, NC, pp. 127-134.
- Baujard, P., Duncan, L.W., Pariselle, A. and Sarr, E., 1987. Étude des effets de quatre nématicides fumigants sur les nématodes et l'arachide au Sénégal. *Rev. Nématol.*, 10: 355-360.
- Bernard, E.C. and Hussey, R.S., 1979. Population development and effects of the spiral nematode *Helicotylenchus dihystera* on cotton in microplots. *Plant Dis. Rep.*, 63: 807-810.
- Byrd, D.W., Jr., Barker, K.R., Ferris, H., Nusbaum, C.J., Griffin, W.E., Small, R.H. and Stone, C.A., 1976. Two semi-automatic elutriators for extracting nematodes and certain fungi from soil. *J. Nematol.*, 8: 206-212.
- Chambers, J.M., Cleveland, W.S., Kleiner, B. and Tukey, P.A., 1983. *Graphical Methods for Data Analysis*. Duxbury Press, Boston, 395 pp.
- Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.*, 74: 829-836.
- Cleveland, W.S., 1985. *The Elements of Graphing Data*. Wadsworth, Monterey, 305 pp.
- Cleveland, W.S. and Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Stat. Assoc.*, 83: 596-610.
- Duncan, L.W. and Ferris, H., 1983. Validation of a model for prediction of host damage by two nematode species. *J. Nematol.*, 15: 227-234.
- Ferris, H., 1978. Nematode economic thresholds: derivation, requirements and theoretical considerations. *J. Nematol.*, 10: 341-350.
- Ferris, H., 1984. Nematode damage functions: the problems of experimental and sampling error. *J. Nematol.*, 16: 1-9.
- Ferris, H., Mullens, T.A. and Foord, K.E., 1990. Stability and characteristics of spatial description parameters for nematode populations. *J. Nematol.*, 22: 427-439.
- Heath, J., Brown, D.J.F. and Boag, B., 1977. *Provisional Atlas of the Nematodes of the British Isles*. Institute of Terrestrial Ecology, Abbots Ripton, UK, 64 pp.
- Hunt, R., 1982. *Plant Growth Curves: The functional approach to plant growth analysis*. University Park Press, Baltimore, 248 pp.
- Jones, F.G.W., 1956. Soil populations of beet eelworms (*Heterodera schachtii* Schm.) in relation to cropping. II. Microplot and field plot results. *Ann. Appl. Biol.*, 44: 25-26.

- McKenry, M.V. and Kretsch, J., 1987. Survey of nematodes associated with almond production in California. *Plant Dis.*, 71: 71-73.
- Oostenbrink, M., 1966. Major characteristics of the relation between nematodes and plants. *Meded. Landbouwhoges. Wageningen*, 66-4.
- Sasser, J.N., 1972. Nematode disease in cotton. In: J.M. Webster (Editor), *Economic Nematology*. Academic Press, London, pp. 187-214.
- Seinhorst, J.W., 1965. The relation between population density and damage to plants. *Nematologica*, 11: 137-154.
- Snedecor, G.W. and Cochran, W.G., 1968. *Statistical Methods*. Iowa State University Press, Ames, IA, 593 pp.
- Spearman, C., 1904. The proof and measurement of association between two things. *Am. J. Psychol.*, 15: 72-101.
- Stynes, B.A., Wallace, H.R. and Veitch, L.G., 1979. A synoptic approach. I. An appraisal of the physical and chemical soil properties in the study area. *Aust. J. Soil Res.*, 17: 217-225.

REPRINTED FROM:

# ECOLOGICAL MODELLING

International Journal on Ecological Modelling and Systems Ecology

---

*Ecological Modelling*, 69 (1993) 243–265  
Elsevier Science Publishers B.V., Amsterdam

## Masked trends in field data: nematode population assessments as an example

H. Ferris <sup>a</sup> and J.-C. Prot <sup>b</sup>

<sup>a</sup> *Department of Nematology, University of California, Davis, CA, USA*

<sup>b</sup> *Department of Plant Pathology, International Rice Research Institute, P.O. Box 933, 1099  
Manila, Philippines*

(Received 25 October 1991; accepted 26 October 1992)

Fonds Documentaire ORSTOM

Cote : B \* 7825 Ex: 1



ELSEVIER

Amsterdam—London—New York—Tokyo

Fonds Documentaire ORSTOM



010007825